

Explainability for artificial intelligence in healthcare: a multidisciplinary perspective

Year: 2020 | Citations: 1207 | Authors: J. Amann, A. Blasimme, E. Vayena, D. Frey, V. Madai

Abstract

Background Explainability is one of the most heavily debated topics when it comes to the application of artificial intelligence (AI) in healthcare. Even though AI-driven systems have been shown to outperform humans in certain analytical tasks, the lack of explainability continues to spark criticism. Yet, explainability is not a purely technological issue, instead it invokes a host of medical, legal, ethical, and societal questions that require thorough exploration. This paper provides a comprehensive assessment of the role of explainability in medical AI and makes an ethical evaluation of what explainability means for the adoption of AI-driven tools into clinical practice. **Methods** Taking AI-based clinical decision support systems as a case in point, we adopted a multidisciplinary approach to analyze the relevance of explainability for medical AI from the technological, legal, medical, and patient perspectives. Drawing on the findings of this conceptual analysis, we then conducted an ethical assessment using the “Principles of Biomedical Ethics” by Beauchamp and Childress (autonomy, beneficence, nonmaleficence, and justice) as an analytical framework to determine the need for explainability in medical AI. **Results** Each of the domains highlights a different set of core considerations and values that are relevant for understanding the role of explainability in clinical practice. From the technological point of view, explainability has to be considered both in terms how it can be achieved and what is beneficial from a development perspective. When looking at the legal perspective we identified informed consent, certification and approval as medical devices, and liability as core touchpoints for explainability. Both the medical and patient perspectives emphasize the importance of considering the interplay between human actors and medical AI. We conclude that omitting explainability in clinical decision support systems poses a threat to core ethical values in medicine and may have detrimental consequences for individual and public health. **Conclusions** To ensure that medical AI lives up to its promises, there is a need to sensitize developers, healthcare professionals, and legislators to the challenges and limitations of opaque algorithms in medical AI and to foster multidisciplinary collaboration moving forward.